



Customer Churn Prediction Using Ordinary Artificial Neural Network and Convolutional Neural Network Algorithms: A Comparative Performance Assessment

Omer Faruk SEYMEN¹ , Emre OLMEZ² , Onur DOGAN^{3,*} , Orhan ER⁴ , Abdulkadir HIZIROGLU⁵ 

¹ Sakarya University, Department of Information Systems Engineering, 54050, Sakarya, Turkey

² Yozgat Bozok University, Department of Mechatronics Engineering, 66900, Yozgat, Turkey

³ Izmir Bakircay University, Department of Industrial Engineering, 35665, Izmir, Turkey

⁴ Izmir Bakircay University, Department of Computer Engineering, 35665, Izmir, Turkey

⁵ Izmir Bakircay University, Department of Management Information Systems, 35665, Izmir, Turkey

Highlights

- In this study, ANN and CNN methods are used for churn prediction.
- High success rate was achieved in all used metrics with both methods.
- It was seen that CNN is more successful than ANN for the churn prediction dataset.

Article Info

Received: 08 Sep 2021
Accepted: 25 Apr 2022

Keywords

Churn prediction
Convolution neural network
Artificial neural network
Deep learning

Abstract

Churn studies have been used for many years to increase profitability as well as to make customer-company relations sustainable. Ordinary artificial neural network (ANN) and convolution neural network (CNN) are widely used in churn analysis due to their ability to process large amounts of customer data. In this study, an ANN and a CNN model are proposed to predict whether customers in the retail industry will churn in the future. The models we proposed were compared with many machine learning methods that are frequently used in churn prediction studies. The results of the models were compared via accuracy classification tools, which are precision, recall, and AUC. The study results showed that the proposed deep learning-based churn prediction model has a better classification performance. The CNN model produced a 97.62% of accuracy rate which resulted in a better classification and prediction success than other compared models.

1. INTRODUCTION

Customer analytics has become a pervasive buzzword in the industry. It is employed to make critical business decisions to model customer behaviors using predictive analytics. Lifetime value modeling, market/customer segmentation and churn analysis are the most popular topics of customer analytics in the industry and academia. Information obtained from these application domains is used for different purposes such as customer relationship management and direct marketing. Yet, companies have still much to obtain from different business applications of customer analytics [1]. Although companies in the retail sector have a great chance with many customers and a huge customer transaction volume, many retailers cannot benefit from customer analytics, especially from churn analysis [2].

The churn concept is a crucial application area of customer analytics from the widest perspective. Churn analysis can be regarded as a sub-domain of customer behavior modeling in customer analytics. Predicting customer churn could be a complicated considering the difficulties such as having dirty data, ending up a low churn rate, churn event censorship [3]. Customer's transaction patterns and other factors directly or indirectly affecting the customer's choices are mainly used for the prediction. Churn prediction is mainly

*Corresponding author, e-mail: onur.dogan@bakircay.edu.tr

regard as a sort of binary classification problem. However, churn studies can focus on churn rate, and this makes them no longer a binary classification. According to a recent sectoral report on customer-indexed, customer churn rates were 20% in telecommunications, 21% in online retail, 22% in financial/credit, 24% in the general retail, 17% in the travel sector, 11% in the big-box electronics and 25% in the cable sector [4]. The majority of customers lie in the non-churn class. This low-churn rate resulting from imbalanced nature makes it difficult to achieve an actual correct classification rate by traditional machine learning methods [5]. Because of that, machine learning-based classification gives high accuracy rates but rather low precision in terms of predicting the churner's class as a performance indicator. These churn prediction confusions demonstrate that a margin of performance improvement is still needed in this area, keeping possibilities emerging in big data analytics.

A well-organized churn results highly depend on variables chosen from the customer dataset as inputs of the prediction model [6]. The common methods contain two main drawbacks [7]. First, they usually require so much time to extract features from thousands of customer attributes. Besides, an expert is mainly needed to apply these methods. Second, traditional churn models are usually developed and applied for a special dataset. Deep learning (DL) methods could be utilized in overcoming these limitations of traditional churn models. First of all, they are faster than traditional methods. Moreover, DL approaches can determine beneficial features without any prior knowledge and expert support. Therefore, applying a DL method for churn prediction in the retail sector, including many data attributes, can produce high accuracy rates in the prediction results.

Convolutional neural networks (CNNs) are specialized types of neural networks and can be applied to many kinds of data with different dimensions [8]. The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any expert support. This benefit makes CNN more attractive to apply a churn prediction model with many data features. In addition to an ordinary ANN implementation, this study proposes a methodology with CNN structure for churn prediction using customer data, including both transactions and demographic attributes, and evaluates performances considering other churn modeling approaches, ordinary neural network, and various machine learning techniques. CNN is actually a method used for image processing. This study's main contribution to the literature is that it applies CNN by transforming the customer transactions and demographic data, which are in matrix form, into an image. The experimental results showed that the CNN-based prediction method produces better accuracy rates than traditional models without manual feature extraction.

The study's remaining part is organized as follows: Section 2 briefly presents related work about churn prediction and some previous studies applied in similar domains. Section 3 gives the background of the CNN. The details of the applied methodology are given in Section 4. The results of the study and comparison with the other methods are shown in Section 5. Finally, conclusion part is presented in Section 6.

2. LITERATURE REVIEW

Customer loyalty and customer churn topics are often discussed in customer analytics studies conducted within the scope of CRM. These studies cover models including data mining, machine learning, and deep learning methods, as well as comparisons of these models with each other. In the literature, customer churn studies performed by these methods are mostly used in the telecommunications and finance sectors where a written agreement is required for customers to start taking service with the company and then terminating the service. For this reason, it is known which customers are missing, and which ones continue to receive service. Churn studies can be carried out with the help of past information and other information of customers. In some sectors such as retail, online retail, or travel, since there is no written agreement between the company and the customer, it is not easy to follow these customers, as the customer cannot be prevented from choosing the company from which they will receive service. Thus, large retail chains offer loyalty card services to customers, gathering information for customer analytics.

In retail, if a customer does not shop for three consecutive months, it is considered as a lost customer [9-11]. In this study, the mentioned timestamp was used to identify lost customers in the data. In recent years,

deep learning, and its types such as CNN and Long Short-Term Memory (LSTM) have also started to take their place in churn studies. Deep learning methods have more powerful learning ability, better prediction accuracy and combine feature extraction and feature selection together in a single architecture [12].

Table 1 summarizes the related work. The bold and first-written performance measurements indicate the metrics that the best score was obtained in the last column of the table. As mentioned before, customer churn studies involving deep learning models are mostly carried out in the telecommunications sector, where customers and company have a mutual contract. [7] developed a Feedforward Neural Network model, including small and large versions, and compared it with ANN and CNN models. By using two different telecom datasets, they compared all models with accuracy metric. [13] also used deep learning methods CNN and LSTM and developed a RFM based daily churn prediction model instead of a monthly churn prediction. The data used in the study was 150 days long telecom data with 5 million observations. CNN and LSTM models were compared with Area Under Curve (AUC), F1 Score, Lift, Log Loss, and Expected Maximum Profit Measure for Churn (EPMC). Results showed that LSTM was slightly better than CNN models while both LSTM and CNN outperformed RFM-based models. [14] used the CNN model to transform the call details, SMS, Recharge, and Data usage records into images to predict churn. Precision, recall, and F1-scores were used to compare the CNN model with the machine learning algorithms, Support Vector Machines, Random Forest, and Gradient Boosting Classifier. CNN model also used by [12] had developed a Natural Language Processing unit that analyzes customer's call transcripts to predict churn. This model conducts a text mining process to reveal churn signals from telecom customers' call records. Precision, recall, and F1-score were used for the evaluation of CNN and NLP models.

Tariq et al. [15], Lalwani et al. [16], and Garimella et al. [17] used the same dataset to predict the churn of telecom customers. [15] used CNN, and [17] used Deep CNN. In contrast, [16] used several machine learning techniques such as AdaBoost, Support Vector Machine (SVM), Logistic Regression (LR), Naïve Bayes (NB), Random Forest (RF), Decision Trees (DT), and XGBoost. These researchers used Confusion Matrix outputs accuracy, AUC, and loss margin. [17] used the maximum dice coefficient and the Jaccard coefficient as performance indicators along with its accuracy. [15] and [17] did not compare their models with any other model; they only presented the performance results. [16], on the other hand, compared all the machine learning techniques and found AdaBoost had the best AUC and accuracy. [18] used a Vector Embedding Model for loss estimation for a telecom dataset of 3333 customers and did not compare the proposed model with any other model. They only presented the accuracy and F1-score of the model proposed, which indicate that the model distinguishes well between churn and non-churn customers.

De Caigny et al. [19] and Evermann et al. [20] developed CNN churn prediction models in the financial service domain. [20] used the Recurrent Neural Network model and Long-Short Term Memory model to predict financial service provider's customers and compared the two models with Precision metric. [19] proposed a prediction model which combines the textual data and structural data of customers in CNN. Emails, contacts, and online connections logs were used for text mining while customer demographics and client behaviors were used as structured data. The authors used AUC and Top Decile Lift (TDL) to compare the performances of CNN-based models, in which one model uses only structured data, one model uses only textual data, and the last one combines both types of data. [21] proposed a Deep Neural Network model for predicting churn of the 10,000 banking customers and compared the model with the Multiplayer Perceptron (MPL) using accuracy and loss margin. For MPL and DNN models the authors also examined the effects of several combinations of activation functions in both hidden and output layers and evaluated several training algorithms such as AdaGrad, Adadelta, AdaMax etc. DNN was superior to MPL on both activation functions and training algorithms.

Unlu [22] proposed an SVM model that uses Bayesian Optimization to optimize the kernels of the SVM. The authors used Linear, Polynomial, Radial, and Sigmoid kernels in SVM models and compared the performances with accuracy, precision, recall, and F1-score. SVM with Linear Kernel was found to be the best to predict churn behaviors of credit card customers. Ozmen and Ozcan [23] developed an Extended Convolutional Decision Tree (ECDT) model to predict the churn of 1186 retail employees. The proposed model was developed by applying Grid Search Optimization (GRID) to improve the classification accuracy of ECDT. The accuracy, precision, recall, and F-measure scores were used for comparison of the proposed

model along with, ECDT, KNN, CNN, SVM, NB, and DT. ECDT-GRID performed 11% better classification than DT, which was the best classifier in basic machine learning algorithms.

Researchers also conducted churn studies in online gaming industry. CNN and LSTM models [24], LSTM and LSTM-based models [25], and LSTM and other machine learning models [26] were used in these churn studies, which analyze the players' in-game purchases, game logs, and membership status. Accuracy, AUC, and F1-score were used as evaluation metrics for these models.

In some studies researchers embedded new features into baseline CNN architecture to predict churn and made comparisons. [27] created a shallow based CNN model to combine the strengths of the deep learning and the shallow model. [28] added the output of LSTM as a new input for the CNN model and [8] used the static data (demographics) as a new feature input into dynamic features (RFM and length of relationship) of CNN. The proposed models in these studies compared with baseline CNN, LSTM, and other shallow machine learning tools with some evaluation metrics such as AUC, Accuracy, Sensitivity, Recall, and F1-score.

In the literature, while there are studies on data mining and machine learning in the retail domain, there are very few studies involving deep learning models. [29] reached out to people who bought new smartphones via a questionnaire on the internet, and their opinions were asked on issues such as the brand, the model, and satisfaction of the phone they purchased. Two years later, when the same people bought a new phone, they asked them what model of smartphone they bought, and they accepted different brand buyers as lost customers. LSTM architecture was used in the study and the results were evaluated with AUC, Precision, and Recall. Another study conducted by [30] compared two deep learning algorithms to analyze historical behavioral patterns of customers with CNN and Restricted Boltzmann Machine to predict churn. The performances were evaluated with Sensitivity, Specificity, Accuracy, Precision, Recall, F1-score, Positive Predicted Value, and Negative Predicted Value.

Table 1. Comparative models and techniques

Cite	Domain	ANN	CNN	LSTM	RNN	SVM	RF	LR	GBC	RBM	OML	Data	Evaluation Metrics	Best
[7]	Telecom	+	+									Two datasets; 71,047 and 3,333 customers	A	93.10%
[12]	Telecom		+									20,000 call transcripts	R, F, P	85.18%
[13]	Telecom		+	+							+	150 days, 5 million data	AUC, F, L, LL, EMPC	91.40%
[14]	Telecom		+			+	+		+			60 days; 18,000 customers' call, SMS and recharge data	R, AUC, F, P	95.00%
[20]	Finance			+	+							13,087 financial customers	P	94.20%
[19]	Finance		+									12-month, 607,125 customers	AUC, L	89.875%
[24]	Online Game		+	+			+	+	+			Three datasets; (16-6-7) months; 193,443 online players	AUC	82.40%
[25]	Online Game	+		+								7 months; 814,822 online players	AUC, A, F	87.37%
[26]	Online Game			+							+	5 months; 10,000 online players	A, F	75.13%
[27]	Insurance	+	+	+			+				+	15 million data	AUC, A, F, P, R	93.35%
[8]	Internet fund	+	+			+	+	+			+	10 months; 41,000 customers	AUC, A, L	94.10%
[28]	Online music streaming		+	+			+				+	1,118,165 customers	AUC, P, R	87.03%
[29]	Retail			+								2 years; 74,088 customers	R, A, AUC, F, P	95.13%
[30]	Retail		+								+	No data information	R, A, F, P	93.00%
This Study	Retail	+	+									27 months old data; 5747 customers; 567 features	P, R, A, AUC	98.27%

A: Accuracy, AUC: Area under curve, F: F1 score, L: Lift, LL: Log Loss, EMPC: Expected Maximum Profit Measure for Churn, P: Precision, R: Recall, Sen: Sensitivity, Spe: Specificity

3. CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks differ from ordinary ANNs with their 2- or 3-dimensional inputs and shared weight architectures. With these features, CNN emerges as a suitable method for image processing problems [31, 32]. Recently, CNNs are frequently used in data processing problems with a large number of features, as well as in image processing problems [13-19].

As seen in Figure 1, in the fully connected architecture of ordinary ANNs, each neuron is connected to all neurons in the next layer with different weights. In the convolution layer of CNN, a convolution matrix moves over the input image and applies the convolution process to obtain the values of the next layer. This architecture is called shared weight architecture because the different regions on the input image are processed with the same weights. Shared weight architecture prevents over-fitting of the designed model to the training set by significantly reducing the number of parameters to be trained in CNN compared to ANN, and this increases the success of generalization.

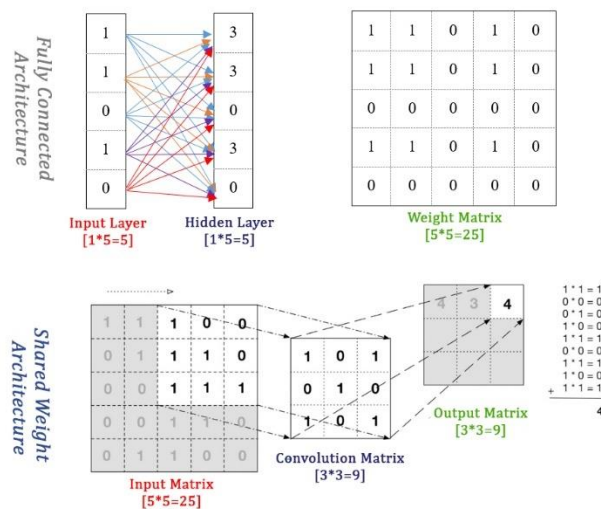


Figure 1. Fully connected vs shared weight architectures

The ReLU activation function is favored in CNN applications over activation functions such as sigmoid and tanh because it is easy to derivatize and this avoids the vanishing gradient problem.

A convolutional neural network can perform the related classification task with high accuracy if the parameters are set to appropriate values. In fact, ANN training is to find the appropriate parameter values (weights) for the relevant task. There are two basic concepts for ANN training, performance measurement and finding parameter values that maximize performance. A cost function used in ANN training calculates the difference of the result produced by ANN from the actual result. An optimization algorithm is used also finds the parameter values that minimize the cost function.

The cross-entropy loss function seen in Equation (1) is frequently used in ANN training. For each input sample, this function calculates how much the value predicted by ANN differs from the actual value. In the cross entropy loss function the logarithm function ensures that the error gets smaller as the value produced by ANN increases for the class to which the input belongs. For classes to which the input does not belong, the exact opposite is true [20, 33]

$$E(\theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n \left[y_k^{(i)} \log(\hat{y}_k^{(i)}) + (1 - y_k^{(i)}) \log(1 - \hat{y}_k^{(i)}) \right] \quad (1)$$

where m is the number of training example, k is the number of classes. If i th sample belongs to class k , $y_k^{(i)} = 1$, else $y_k^{(i)} = 0$. $\hat{y}_k^{(i)}$ is the predicted value of the class k for the sample i .

ANN training is actually finding the parameter values that make the cost function minimum. The gradient descent algorithm seen in Equation (2) multiplies the gradient of the cost function with a certain learning rate and subtracts it from the current parameter values. To find the appropriate parameter values this process is continued iteratively until the gradient values are sufficiently small

$$\theta_{i+1} = \theta_i - \alpha \nabla E(\theta_i) \quad (2)$$

where θ indicates the parameter vector, i is the iteration number, α shows the learning rate, ∇ refers to the gradient operator.

4. METHODOLOGY

In this section, data preparation and framework of the study were presented. As mentioned in the literature review, in the retail sector, customers are considered lost if they do not shop for three consecutive months. So, if a customer has no shopping data three months or more, he/she is considered as a churn customer while others labeled as non-churn customers. After knowing which customers were churn, only 24 months of data were used to train and test the datasets. The dataset (5747) is randomly divided into 70% (4023) for training and 30% (1724) for testing as known traditional validation approach.

4.1. Data Preparation and Study Framework

The pertaining data descriptions, basic statistics about the dataset and demographics of customers are shown in Tables 2, 3 and 4. The retail data obtained from a national supermarket chain covers 27 months old retail scanner data of 5747 customers between January 2018 and March 2020. Retail data includes 593 features such as the customer's identification number, spending per product category, demographics information, promotion in product categories by month, and number of purchasing per product category when the promotion is applied. Since the company collects the products under eight basic product categories, it is used in this study as well.

Table 2. Basic statistics of the dataset

Data Source	Retail Scanner Data
Number of Observations	5,747
Total Number of Product Categories	8
Total Number of Independent Variables	593
Total Number of Churn Customers	2507
Churn Customers Percentage	43,62%
Total Number of Nonchurn Customers	3240
Nonchurn Customers Percentage	56,38%
Total Number of Training Data	4023
Total Number of Test Data	1724

Table 3. Demographics of customers

Attributes	Categories	Men		Women	
		Frequency	Percent	Frequency	Percent
Customers Size	5747	3219	56,01%	2528	43,99%
Age (years old)	30 years and below	31	0,5%	27	0,5%
	31 – 40 years	651	11%	564	10%
	41 – 50 years	1050	18%	852	15%
	51 – 60 years	682	12%	453	8%
	More than 61 years	805	14%	632	11%
Education	High School and below	1306	22,7%	1096	19,1%
	Associate Degree 3	1001	17,5%	875	15,2%
	Graduate 4	668	11,6%	439	7,6%
	Post Graduate 5	244	4,3%	118	2%

Income (annual)	Less than 30,000 TL	12	206	3,6%	443	7,7%
	30,001 – 60,000 TL	345	1101	19,2%	1097	19,1%
	60,001 – 90,000 TL	678	1155	20,1%	662	11,5%
	90,001 – 120,000 TL	678	575	10%	243	4,2%
	More than 120,000 TL	678	182	3,2%	83	1,4%
Working Status	Employed Full Time		2455	42,7%	1684	29,3%
	Employed Part Time		158	2,8%	330	5,7%
	Unemployed		137	2,4%	119	2,1%
	Retired		469	8,1%	395	6,9%

Table 4. Data description

Variable Type	Variable	Feature Size
Identification	ID number of loyalty cards given by the company and representing customers	1 (Not included in prediction model)
Behavioral	Spending per product category for every time interval (total spending per product category for each month)	24 months and 8 product categories = 192
	Promotions in product categories by month (Information on whether there is a promotion per product categories by month)	24 months and 8 product categories = 192
	Number of purchasing per product category under the promotion (When a promotion is made in a product category, whether the customer purchased from that product category)	24 months and 8 product categories = 192
Demographics	Age, Sex, Marital Status, Education, Job, Resident Status, Having Children, Spouse Education, Job, Possession of some devices	17

Customers with missing values were removed from the data in the preprocessing step. In the model building section, the data was divided into training and test datasets for all models. Along with the CNN and ANN deep learning models, other machine learning models also were used for the prediction of the accuracy of churn detection. These models are Logistic Regression, Support Vector Machines (Linear, Cubic, Fine Gaussian, Medium Gaussian, Coarse Gaussian), and k-nearest neighborhood-KNN (Fine, Medium, Coarse, Cosine, Cubic, Weighted) models.

The final step of the model is to compare the proposed CNN model with ANN and the other machine learning models by using performance metrics such as Area Under Curve, Precision, Accuracy, and Recall. Figure 2 shows the framework of the study.

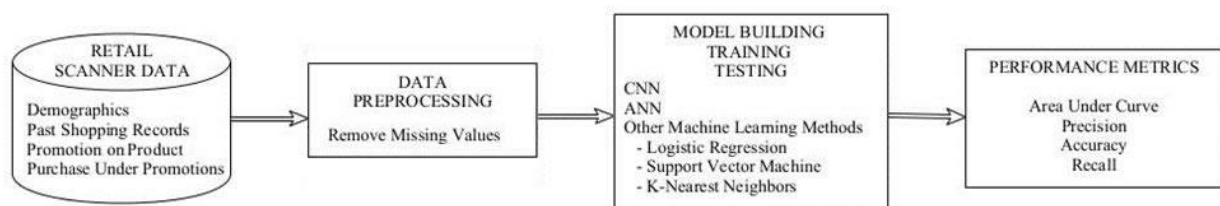


Figure 2. Framework of the study

4.2. Proposed ANN and CNN Models

In this study, among the ANN models with different layers and neuron numbers, the model that made the most successful classification (%96) for the churn data set was chosen as the reference ANN model. The proposed ANN architecture has three hidden layers, each consisting of 20 neurons, as shown in Figure 3. The CNN model, which is used for performance comparison with the ANN model, was designed by adding a convolution layer which consist of 30 filters of 3x3 size to the input of the reference ANN model. Customer demographics, purchase under promotion data, promotion periods and spending data were used as inputs for the CNN model. While the data set consisting of 567 features can be applied as vector input to the ANN model, the data is converted into a 21x27 matrix form to apply it to the CNN model. To speed up training and obtain higher learning rates, batch norm process was used in convolution layer. The ReLU function, which converts negative values to zero while using positive values as direct input, was also used. ReLU function is used to increase non-linearity to support network learning. The output of convolution layer was flattened to 14250×1 and linked to the three fully connected layers by $20 \times 20 \times 20$. The designed CNN architecture is shown in Figure 4. For the training of the designed ANN and CNN models, Adam optimization algorithm was used with a learning rate of 0.001. For each model, the epoch number was set to 10 and the mini-batch size was set to 128.

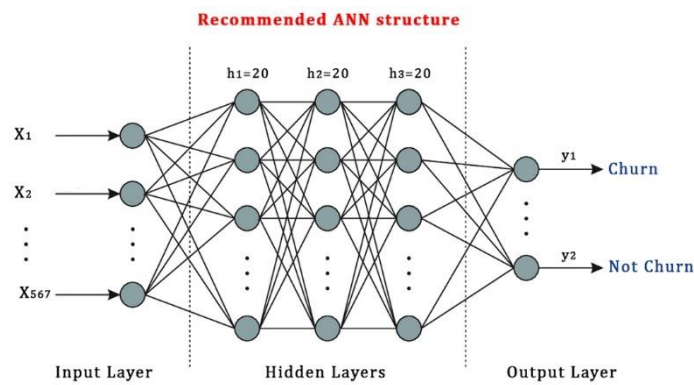


Figure 3. Proposed ANN architecture for churn prediction

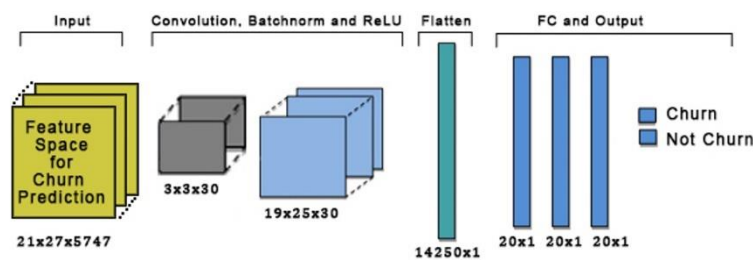


Figure 4. Proposed CNN structure for churn prediction

5. RESULTS

In this section, ANN and CNN results based on the classification accuracy were evaluated and compared to predict customer churn. All numerical results were obtained by using MATLAB R2021a on a core i7, 2.7 GHz processor, 16GB RAM under Windows 10 operating system. The traditional validation approach was used to evaluate the performance of the proposed algorithms. Two different kinds of classification algorithms were applied, and their performance scores are presented in Table 5.

Table 5. Performance results of ANN and CNN methods

Algorithms	Accuracy	Precision	Recall	AUC
ANN	0.96	0.9827	0.93	0.963
CNN	0.9762	0.9747	0.971	0.976

According to the results the performance scores of the CNN structure are higher than the ANN structure. However, the difference between them is not too much. It is understood that the performances of the two methods are close in customer churn prediction problems. Moreover, in this problem, the number of input features is excessive for an ordinary ANN structure and therefore increases the processing time. CNN has a significant advantage because of the feature extraction layer. Therefore, CNN structure is suggested for problems with large input features. The confusion matrix (Figure 5) and ROC curve (Figure 6) obtained from the training and testing of the ANN and CNN algorithms supports this suggestion.

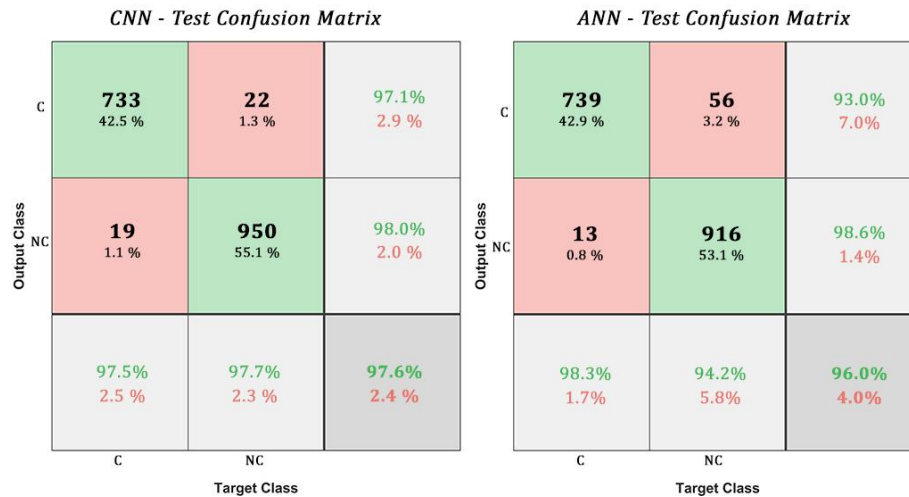


Figure 5. Confusion matrices of CNN and ANN structures

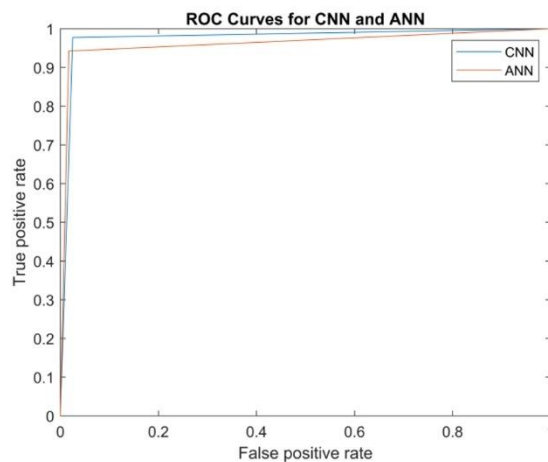


Figure 6. ROC curves for CNN and ANN

The confusion matrix comparing the predictions and actual values of the target attribute is given in Figure 5. As seen in the CNN confusion matrix, 1683 (733+950) of 1724 samples were classified as true and 41 (22+19) as false. In the ANN confusion matrix, 1655 (739+916) of 1724 samples were true classified and 69 (56+13) as false.

ROC curve (A receiver operating characteristic curve) is a graphical plot showing the ability of a binary classifier system to classify as the discrimination threshold changes. The ROC curve obtained as a result of determining the ANN and CNN performances is given in Figure 6.

Customer data were also tested with various machine learning techniques to compare the study's success, and results were given in Table 6. The performance of the CNN algorithm is more successful than all machine learning methods that were selected for comparison.

Table 6. Comparison with selected machine learning techniques

Machine Learning Models	Accuracy
Logistic Regression	93.3
Linear SVM	95.9
Cubic SVM	96
Fine Gaussian SVM	65.5
Medium Gaussian SVM	94.7
Coarse Gaussian SVM	93
Fine KNN	73.5
Medium KNN	70.5
Coarse KNN	67.8
Cosine KNN	83.9
Cubic KNN	71.8
Weighted KNN	73.5

6. CONCLUSION

This paper presents a comparative study for customer churn prediction using ordinary ANN and CNN algorithms based on a dataset obtained from a national supermarket's 27-month retail scanner. At the same time, the results obtained were compared with the realization of different machine learning techniques.

When the results obtained are examined, it was understood that the ANN and CNN methods mentioned above have an acceptable success for the prediction of customer churn. Accuracy performances could be higher if the number of samples in the dataset is increased.

Two significant results of the study can be summarized. First, The CNN architecture proposed in the study gave the best performance on the customer churn dataset. However, the performance of the ANN architecture is also very close to CNN. Therefore, both methods can be preferred interchangeably for this problem. Second, different machine learning methods were performed to compare the results. The classification performances vary from 70% to 96%. One of the machine learning methods, the Cubic SVM method, has a similar accuracy score with ANN. CNN can be considered as the most effective method to predict customer churn based on the domain and the dataset selected with such a large number of input features.

Artificial neural networks and deep learning algorithms have been successful in predicting customer churn. It is seen that deep learning techniques give better results in more complex structures. Although the ANN and CNN models used in this study have a very high accuracy rate, they do not indicate how much each input feature impacts the output. In future studies, different ML algorithms can be used to determine the importance of input features in order to determine the marketing strategy as well as to increase the success rate from a practical point of view. In addition, it is predicted that higher success rates will be achieved with the improvement of deep learning techniques over time.

The performance of other predictive models can be explored in future studies as they are developed, and a combination of these methods can also be used. In addition, web-based tools can be created by which companies can make predictions with machine learning-based estimation software. Future studies can also be extended to explore changing behavior patterns of lost customers by applying artificial intelligence techniques for forecasting and trend analysis.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

- [1] Germann, F., Lilien, G., Moorman, and C., Fiedler, L., “Driving customer analytics from the top. Customer Needs and Solutions”, *Customer Needs and Solutions*, 7(3):43–61, (2021).
- [2] Germann, F, Lilien, G.L, Fiedler, L, and Kraus, M., “Do retailers benefit from deploying customer analytics?”, *Journal of Retailing*, 90(4):587-593, (2014).
- [3] Mishra, A. and Reddy, U. S., “A novel approach for churn prediction using deep learning”, *Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Coimbatore India, 1–4, (2017).
- [4] Aspect Software. Aspect Consumer Index 2020. <https://www.aspect.com/the-aspect-consumer-experience-index-2020>, 2020. Online, accessed 10 April 2021.
- [5] Uzair, A., Khan, A., Saddam, H. K., Basit, A., Haq, I., and Lee, Y. S., “Transfer learning and meta classification based deep churn prediction system for telecom industry”, *arXiv preprint arXiv:1901.06091*, (2019).
- [6] Ladyzynski, P., Zbikowski, K. and Gawrysiak, P., “Direct marketing campaigns in retail banking with the use of deep learning and random forests”, *Expert Systems with Applications*, 134: 28–35, (2019).
- [7] Umayaparvathi, V. and Iyakutti, K., “Automated feature selection and churn prediction using deep learning models”, *International Research Journal of Engineering and Technology (IRJET)*, 4(3): 1846–1854, (2017).
- [8] Wang, C., Han, D., Fan, W. and Liu, Q., “Customer churn prediction with feature embedded convolutional neural network: An empirical study in the internet funds industry”, *International Journal of Computational Intelligence and Applications*, 18(1): 1950003, (2019).
- [9] Buckinx, W. and Poel, D. V., “Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting”, *European Journal of Operational Research*, 164(1): 252–268, (2005).
- [10] Keiningham, T. L., Cooil, B., Aksoy, L., Tor, W. A., and Weiner, J., “The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share-of-wallet”, *Managing Service Quality: An International Journal*, 17(4): 361–384, (2007).
- [11] Migueis, V. L., Poel, D. V., Camanho, A., and Cunha, J. F., “Modeling partial customer churn: On the value of first product-category purchase sequences”, *Expert Systems with Applications*, 39(12): 11250–11256, (2012).
- [12] Zhong, J. and Li, W., “Predicting customer churn in the telecommunication industry by analyzing phone call transcripts with convolutional neural networks”, In *Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence*, Suzhou China, 55–59, (2019).
- [13] Alboukaey, N., Joukhadar, A., and Ghneim, N., “Dynamic behavior-based churn prediction in mobile telecom”, *Expert Systems with Applications*, 162: 113779, 2020.
- [14] Chouiekh, A. and Haj, E. H., “Deep convolutional neural networks for customer churn prediction analysis”, *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 14(1): 1–16, (2020).

- [15] Tariq, M. U., Babar, M., Poulin, M. and Khattak, A. S., “Distributed model for customer churn prediction using convolutional neural network”, *Journal of Modelling in Management*, ahead-of-print, (2021).
- [16] Lalwani, P., Mishra, M. K., Chadha, J. S. and Sethi, P., “Customer churn prediction system: A machine learning approach”, *Computing*, 104(2), 271–294, (2022).
- [17] Garimella, B., Prasad, G. V. S. N. R. V. and Prasad, M. H. M. K., “Churn prediction using optimized deep learning classifier on huge telecom data”, *Journal of Ambient Intelligence and Humanized Computing*, (2021).
- [18] Cenggoro, T. W., Wirastari, R. A., Rudianto, E., Mohadi, M. I., Ratj, D., and Pardamean, B., “Deep Learning as a Vector Embedding Model for Customer Churn”, *Procedia Computer Science*, 179, 624–631, (2021).
- [19] Caigny, A. D., Coussement, K., Bock, K. W., and Lessmann, S., “Incorporating textual information in customer churn prediction models based on a convolutional neural network”, *International Journal of Forecasting*, 36(4): 1563–1578, (2020).
- [20] Evermann, J., Rehse, J. R., and Fettke, P., “Predicting process behaviour using deep learning”, *Decision Support Systems*, 100: 129–140, (2017).
- [21] Domingos, E., Ojeme, B., and Daramola, O., “Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector”, *Computation*, 9(3), 34, (2021).
- [22] Unlu, K. D., “Predicting credit card customer churn using support vector machine based on Bayesian optimization”, *Communications Faculty of Science University of Ankara Series A1Mathematics and Statistics*, 70(2), 827–836, (2021).
- [23] Ozmen, E. P., and Ozcan, T., “A novel deep learning model based on convolutional neural networks for employee churn prediction”, *Journal of Forecasting*, 41(3), 539–550, (2022).
- [24] Kim, S. D. Choi, Lee, E., and Rhee, W., “Churn prediction of mobile and online casual games using play log data”, *PloS One*, 12(7): e0180735, 2017.
- [25] Kristensen, J. T., and Burelli, P., “Combining sequential and aggregated data for churn prediction in casual freemium games”, *2019 IEEE Conference on Games (CoG)*, London England, 1–8, (2019).
- [26] Guitart, A., Chen, P. P., and Perianez, A., “The winning solution to the IEEE CIG 2017 game data mining competition”, *Machine Learning and Knowledge Extraction*, 1(1): 252–264, (2019).
- [27] Zhang, R., Li, W., Tan, W., and Mo, T., “Deep and shallow model for insurance churn prediction service”, In *2017 IEEE International Conference on Services Computing (SCC)*, 346–353, IEEE, (2017).
- [28] Zhou, J., Yan, J., Yang, L., Wang, M., and Xia, P., “Customer churn prediction model based on LSTM and CNN in music streaming”, *DES tech Transactions on Engineering and Technology Research*, (AEMCE), (2019).
- [29] Kim, J., Ji, H. G., Oh, S., Hwang, S., Park, E., and Pobil, A. P., “A deep hybrid learning model for customer repurchase behavior”, *Journal of Retailing and Consumer Services*, 59:102381, (2021).

- [30] Dingli, A., Marmara, V., and Fournier, N. S., “Comparison of deep learning algorithms to predict customer churn within a local retail industry”. *International Journal of Machine Learning and Computing*, 7(5): 128–132, (2017).
- [31] Aghdam, H. H., and Heravi, E. J., *Guide to convolutional neural networks: A Practical Application to Traffic-Sign Detection and Classification*. Springer, New York, NY, 10: 978–973, (2017).
- [32] Lu, L., Zheng, Y., Carneiro, G., and Yang, L., “Deep learning and convolutional neural networks for medical image computing”. *Advances in Computer Vision and Pattern Recognition*, 10: 978–3, (2017).
- [33] Geron, A., “Hands-on machine learning with Scikit-Learn”, *Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*, Second Edition, O’Reilly Media, (2019).